ARMY RESEARCH LABORATORY

Estimating the Parameters of Chaotic Maps

by Scott Hayes

ARL-TR-95 February 1993

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1993	Final, from Jan 92–April 92

4. TITLE AND SUBTITLE
Estimating the Parameters of Chaotic Maps

5. FUNDING NUMBERS
PE: 61102

6. AUTHOR(S)
Scott Hayes

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Attn: AMSRL-WT-NH
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-95

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
AMS code: 611102.H440011
ARL project: 2AE127

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The lowest mean square error (LMSE) parameter estimator for a chaotic map (a discrete-time dynamical system) is developed, its implementation is discussed, and the results of computer simulations evaluating its performance are presented. The estimator detects the control parameter of a one-dimensional chaotic map from observations of the output sequence in the presence of additive noise. This type of estimation problem is of interest for finding dynamical models of physical systems from measurements by estimating the return map for a Poincaré surface of section in state space. Several common estimators for maps are first described. The Cramer-Rao LMSE bound for estimators of chaotic maps is then computed, and the optimal estimator is developed. It is shown that this estimator is efficient (attains the Cramer-Rao bound) and it converges exponentially to the correct parameter for certain trajectories. Because the estimator becomes impractical to use for very long data sequences, an alternative implementation for long data blocks is also described.

The exponentially increasing accuracy of this estimator as a function of sample size indicates that the estimator extracts parameter information at a constant rate, unlike a typical estimator with inverse square root convergence, which has a decaying information rate. The information rate is shown to be given by a parameter entropy function formally similar to the Kolmogorov entropy for maps. The information rate is therefore quantitatively linked to the global chaos of the system through this parameter entropy.

14. SUBJECT TERMS
Parameter, estimation, map, chaotic, dynamical system

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Contents

Figures

1. Introduction

Estimators for the equations of motion of dynamical systems are of interest for a number of applications, including control of chaos [1] and model identification for signal processing. Estimation of the differential equations of motion for a continuous-time system is complicated by a number of factors, even if the topology of the system is well understood. These factors include the availability of only undersampled and noisy data, and the need for high-order basis expansions to fit the differential equations in a coordinate delay embedding. Fortunately, a sufficient description of a dynamical system can often be made using only low-dimensional discrete-time return maps. This is the case with the highly successful techniques recently developed for the control and targeting of chaotic dynamical systems [1,2]. The return map is usually estimated with the use of bin averaging, or function fitting, which yields sufficient accuracy for many purposes. For technological applications, however, the ability to quickly and accurately estimate return maps in the presence of noise with drifting or intentionally altered system parameters becomes important. A technique for doing this for a one-dimensional chaotic map with known functional form is considered here. This estimation technique can be regarded as coherent parameter detection, as opposed to other methods which do not look at long-range correlations in the data sequence. The application of the technique is demonstrated using the logistic map, and it is shown that the estimator is efficient for trajectories where an efficient estimator exists. The technique should be extendible to higher-dimensional systems, without major complications.

To conclude the report, a more general result is considered: the interpretation of a certain entropy-like quantity as a parameter information acquisition rate. This result is obtained by considering the exponential convergence rate of the efficient estimator. Thus, the detection of the map parameter using the efficient estimation technique can be thought of as obtaining parameter information at a finite rate.

2. Discrete-Time Chaotic Dynamical Systems

In a discrete-time dynamical system, the time parameter takes on integer values. The system is assumed to have definable and distinct configurations or states that can be represented by a state point or vector in a state space. Mathematically, a discrete-time system is usually described by a mapping of previous state-space vectors into the next one, with time as an integer index:

$$\mathbf{X}_{n+1} = \mathbf{F}(\mathbf{X}_n, \mathbf{X}_{n-1}, \ldots, \mathbf{X}_0) \quad . \tag{1}$$

Examples of discrete-time systems are the difference equations of linear system theory, and iterated functions mapping a space into itself. A first-order discrete-time system is one whose current state vector depends only on the previous one. A first-order system with *system noise* present is described by the stochastic map

$$\mathbf{X}_{n+1} = \mathbf{F}\mathbf{X}_n + \boldsymbol{\xi}_{n+1} \ , \tag{2}$$

where $\boldsymbol{\xi}$ is a random vector. Each noise vector is therefore propagated forward in time under the action of the map. A first order system with only *additive noise* present is described by

$$\mathbf{X}_{n+1} = \mathbf{F}(\mathbf{X}_n - \boldsymbol{\xi}_n) + \boldsymbol{\xi}_{n+1} \ . \tag{3}$$

The noise vector is therefore additive but is not propagated forward by the map. Only additive noise is considered in the derivation and evaluation of the estimators here.

Some maps are chaotic; that is, they display a behavior that is complex and aperiodic. For concreteness, the development in this report concentrates on a classic example of a chaotic map, the logistic map. The logistic map [3], a functional transformation of the unit interval on the real axis into itself, is described by the equation

$$x_{n+1} = \theta x_n (1 - x_n) \ . \tag{4}$$

For some values of the parameter θ, the sequence of points generated by the map is chaotic for almost all initial points. Chaotic sequences have the following properties: Sequences generated starting with two nearby points diverge from each other exponentially (called sensitivity to initial conditions), the sequences are aperiodic, and the points are distributed on the real line with a natural probability density $p(x)$. Figure 1 shows the graph of the logistic map for $\theta = 3.7$, along with the tenth iterate of the map. The extreme sensitivity to initial conditions is apparent in the graph of the tenth iterate. Changing x_0 slightly has a large effect on the tenth-iterated value. The global chaos is not dependent on initial conditions; that is, the conditional probability density $p(x|x_0) = p(x)$ for almost all initial points x_0.

The logistic map was investigated in depth by Feigenbaum [4] and shown to be related in many important ways to almost all single-hump maps of an interval of the real axis into itself. In this report, the detailed behavior of the map is not considered. One important aspect of the behavior of the logistic map that is relevant to the parameter estimation problem, however, is the qualitatively different behavior that occurs for different values of the parameter. As the parameter θ is continuously varied from 0 to 4 through the positive real numbers, the nature of the sequence of iterates changes dramatically. Figure 2 is a plot of the points that are generated by the map as a

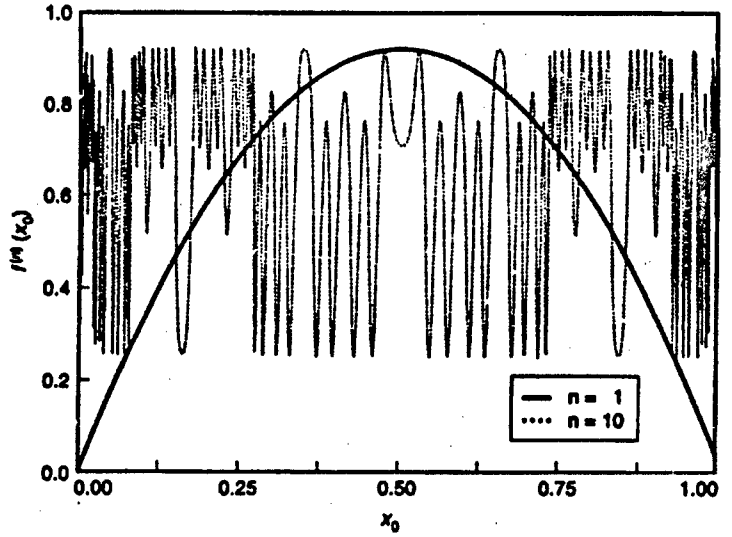


Figure 1. Logistic map for $\theta = 3.7$: first and tenth iterate.

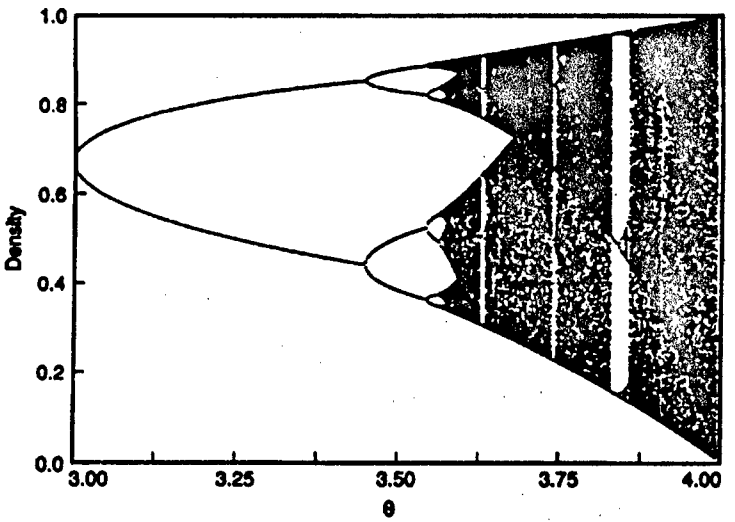


Figure 2. Bifurcation diagram (density versus control parameter) for logistic map.

function of the control parameter θ for 250 iterates. This figure can be viewed as showing the density of the map-generated points on the unit interval versus the control parameter. Darker bands thus correspond to higher density, and lighter regions to lower density. (The map was iterated before plotting to allow the decay of transients. This type of plot is generally called a bifurcation diagram.)

The features of the plot that are important to the estimation problem are as follows: First, as the control parameter is varied from 0 to a critical value of about 3.57, the generated sequence is periodic, with the period length increasing in abrupt jumps as the control parameter is increased. This behavior is called period doubling or bifurcation. Past this critical value, the sequence is aperiodic but still distributed over bands. As θ is increased, the bands move together until the density becomes connected over a single interval. The natural probability density for $\theta = 3.7$ is shown in figure 3. At

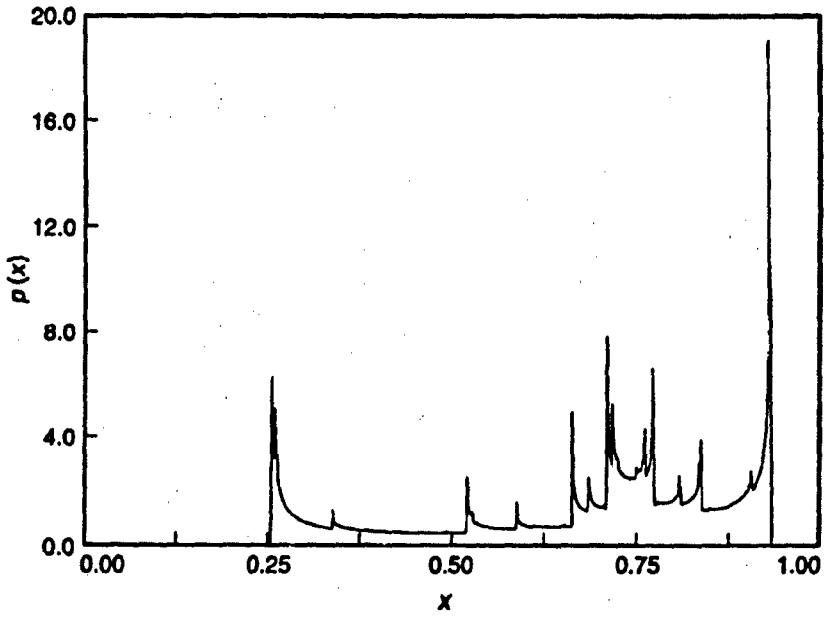


Figure 3. Natural probability density for logistic map at $\theta = 3.7$.

this value of the control parameter, the density is connected over one interval, but there are singularities everywhere. Also important for the parameter estimation problem is the existence of periodic windows in the chaotic regions. The precise distribution of chaotic and periodic parameter values is very complex (the set of chaotic parameter values is a fractal), but the important aspect of this behavior is that the estimation technique that is developed performs best when there is chaos. Because the sensitivity caused by the presence of chaos makes the sequence of iterates generated by the map extremely sensitive to changes in the value of the parameter θ, the parameter can be estimated with great accuracy. If the map is not in the chaotic region, one cannot expect the estimator to perform as well.

3. The Parameter Estimation Problem for Maps

Consider a discrete-time chaotic system described by equation (1), where the map **F** is an implicit function of one or more parameters defined by the parameter vector $\boldsymbol{\theta}$. The problem is to estimate the parameter vector from a finite-time observation of the sequence generated by the system. A parameter estimator for a general map is a function that maps a state vector sequence of a given length into an estimate of the parameters of the map:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Theta}[\mathbf{X}^n] \ . \tag{5}$$

The parameter estimator is a function of the data vector and is therefore itself a random variable. The observations of the sequence can be imprecise because of additive noise or quantization. Also, there may be system noise

present in which each successive state of the system is perturbed by a random variable. The case of system noise is not considered in this development.

Certain attributes of a good parameter estimator are desired. First, the estimator should be nearly unbiased; that is, the expectation of the estimate should equal the correct parameter. Second, the estimator should yield estimates that are consistently close to the correct parameter. Although many possible criteria exist for evaluating the performance of an estimator, the usual measure of performance is the variance of the estimated values. In this report, the usual minimum variance criterion is used, but an important point should be mentioned. Because the treatment given here is for small noise amplitudes and uses local linear approximations to the maps, the impression might be that there exists no efficient estimation procedure for higher noise amplitudes. This is not necessarily so; in fact, it appears likely that the parameter of the map can be extracted efficiently with higher noise levels, but the estimator would have to be more complex. Furthermore, the existence of a threshold of noise amplitude beyond which the optimal estimator considered here deviates from the theoretical performance does not preclude its use in these higher noise levels. The estimator described in this report is therefore optimal only for sufficiently small noise amplitudes, but for uncorrelated noise it will be shown to be the best estimator that can be found in the strict sense that it minimizes the variance of the estimate over the set of all possible estimation functions defined on a given data sequence.

4. Some Parameter Estimators for the Logistic Map

A straightforward way to estimate the parameter of the logistic map is simply to take the i^{th} estimate of the parameter as

$$\hat{\theta}_i = \frac{x_i}{x_{i-1}(1 - x_{i-1})} \quad . \tag{6}$$

This is equivalent to choosing the estimate such that $f(x_{i-1}) = x_i$. The properties of this estimator are now briefly considered. First, the bias in the expectation of the $i = 1$ estimate of θ is given by

$$E\left[\hat{\theta}_1 - \theta\right] = E\left[\frac{\tilde{x}_1 + \xi_1}{(\tilde{x}_0 + \xi_0)(1 - \tilde{x}_0 - \xi_0)} - \frac{\tilde{x}_1}{\tilde{x}_0(1 - \tilde{x}_0)}\right] , \tag{7}$$

where the tildes over the variables indicate that these are the underlying points generated by the map unperturbed by the noise, and ξ_1 is the noise perturbation. (Computing the bias for $i = 1$ is really general for all i and is

done only to simplify the notation.) This expectation can be evaluated for small noise amplitudes by using

$$\left[(\tilde{x}_0+\xi_0)(1-\tilde{x}_0-\xi_0)\right]^{-1} \cong \frac{1}{\tilde{x}_0(1-\tilde{x}_0)} - \frac{(1-2\tilde{x}_0)\xi_0-\xi_0^{\,2}}{\left[\tilde{x}_0(1-\tilde{x}_0)\right]^2} \ . \qquad (8)$$

The bias in the estimate is then given by

$$E[\hat{\theta}_i-\theta] = \frac{\tilde{x}_1}{\left[\tilde{x}_0(1-\tilde{x}_0)\right]^2}\sigma^2 \ , \qquad (9)$$

where σ^2 is the noise variance. The estimator is therefore unbiased to first order. (For the estimator to be biased in the first-order analysis, the bias would have to be directly proportional to the first or lower power of σ.)

These estimates for the parameter are generally combined in some manner to converge to the correct value as the sample size is increased; some algorithms just add them, but it is obvious that the variance of each individual estimate is dependent upon the state of the map. A more accurate estimate is therefore obtained by using

$$\hat{\theta} = \sum_i w_i \hat{\theta}_i \ , \qquad (10)$$

where the w_i form a weight vector to minimize the variance of the estimate. The coefficients are then to be determined. A first-order approach to choosing the w_i approximates the error in the i^{th} estimate of θ as

$$\delta\hat{\theta}_i \propto \frac{1}{x_{i-1}(1-x_{i-1})} \qquad (11)$$

which is reasonable considering the functional form of the map. The minimization of the variance of this estimator using Lagrange multipliers based on the above intuitive expression for the error then yields

$$w_k = \frac{\nu_k^{\,2}}{\sum_i \nu_i^{\,2}} \text{ with } \nu_k = x_{k-1}(1-x_{k-1}) \ . \qquad (12)$$

(The technique for applying Lagrange multipliers to this type of problem is described later in the development of the optimal estimator.) This estimation technique is in fact equivalent to least-squares fitting a second-order polynomial to an embedding of pairs of points from the data sequence. It is also the same as retaining only the diagonal terms in the covariance matrix for the optimal estimator to be developed here, and using a crude estimate for the variance.

The method by which both map estimators and differential equation estimators are usually implemented is equivalent to the crude estimator described above. This equivalence is important because it shows that the common procedure for estimating equations does not nearly achieve the performance that is possible. An implementation of a singular value decomposition (SVD) algorithm is usually used in the following manner to do least-squares fitting, and is sometimes described as minimizing the forward prediction error for maps (as well as differential equations). The standard implementation of the SVD algorithm is capable of least-squares fitting an n^{th} order polynomial to an embedded data set in function space. In the standard implementation, neighboring pairs of data points are embedded in a Cartesian plane so that

$$(X,Y)_i = (x_{i-1}, x_i) \ . \tag{13}$$

This does not reconstruct the attractor of the system as with time delay embedding techniques; it reconstructs the map in function space or, in this case, in the map plane. This reconstruction is illustrated in figure 4, which shows 5000 pairwise embedded points in the map plane for the logistic map with $\theta = 3.7$. Independent Gaussian noise perturbations with a standard deviation of $\sigma = 0.01$ were added to each point in the data sequence produced by the map before the points were paired and embedded. The graph of the unperturbed map function is also shown as a solid curve in the figure. The effects of the noise are evident in the fuzzy picture of the logistic map (an inverted parabola) that is produced. The effects of the attractor and of the singularity distribution of the density are also apparent: The reconstructed function extends only over a portion of the unit interval (only over the

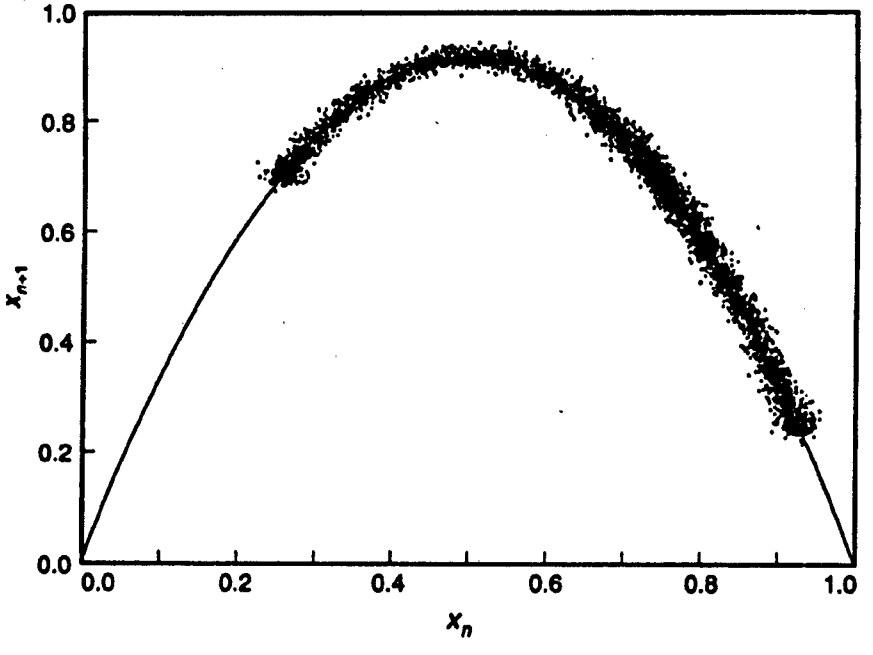


Figure 4. Embedding reconstruction of logistic map for $\theta = 3.7$.

attractor), and there are regions of visibly greater point density near the most prominent singularities. The SVD algorithm is then used to find the polynomial of a given order that fits the embedded points with the lowest mean-square error. For the logistic equation with the SVD algorithm constrained to a second-order polynomial, this is the same as minimizing the sum of the squares of the vertical distances to the estimated logistic map function.

This curve fitting approach will now be shown to be equivalent to the estimator using the weighting of pairwise estimates developed previously. The χ-squared error over all embedded points to be minimized is

$$\chi^2 = \sum_i \left| f^e(x_{i-1};\theta) - x_i \right|^2 \quad . \tag{14}$$

where f^e is the estimated logistic map functional. The condition for minimizing the χ-squared error is thus given by

$$\sum_i \left(f^e(x_{i-1}) - x_i \right) \partial_\theta f^e(x_{i-1}) = 0 \quad . \tag{15}$$

With the use of the expression for the estimated logistic map, $f^e(x_{i-1}) = \hat{\theta} x_{i-1}(1 - x_{i-1})$, and the definition $v_i = x_{i-1}(1 - x_{i-1})$, this condition yields

$$\hat{\theta} = \frac{\sum_i v_i x_i}{\sum_i v_i^2} \tag{16}$$

for the estimated value of the parameter. With $\hat{\theta}_i = (x_i / v_i)$ equation (16) becomes

$$\hat{\theta} = \frac{\sum_i v_i^2 \hat{\theta}_i}{\sum_i v_i^2} \quad . \tag{17}$$

With the value of the weight given in equation (12), equations (17) and (10) are identical. The least-squares fit to the functional form of the map is therefore equivalent to the crude estimator considered previously.

5. The Cramer-Rao Bound for Chaotic Maps

The Cramer-Rao bound [5] for the lowest attainable mean-squared error is now calculated for the logistic map. The calculation shows that the error bound decreases exponentially with the length of the data vector. This is because of the sensitivity of the map to small changes in the parameter value. If a small change in the parameter has a large effect on the data vector, one would expect to be able to estimate the parameter very accurately.

The Cramer-Rao bound for the variance of a single parameter θ is given by [5]

$$\sigma_\theta^2 \geq \frac{1}{E\left[\partial_\theta \log q(\mathbf{X}|\theta)\right]^2} , \tag{18}$$

where the expectation is taken over all possible data vectors **X**, and $q(\mathbf{X}|\theta)$ is the conditional density of the data vector at the parameter value θ. The conditional density for normally distributed noise is dependent on θ only through the mean and is given by

$$q(\mathbf{X}|\theta) = \prod_i \frac{1}{\sqrt{2\pi\sigma}} e^{-(x_i - \mu_i(\theta))^2/2\sigma^2} . \tag{19}$$

Taking the derivative in equation (18) yields

$$\sigma_\theta^2 \geq \frac{1}{E\left[\sum_i \mu_{i,\theta} \frac{x_i - \mu_i}{\sigma^2}\right]^2} , \tag{20}$$

where the subscript ",θ" denotes differentiation with respect to θ. If we assume that the noise vector has uncorrelated components, the expression becomes

$$\sigma_\theta^2 \geq \frac{\sigma^2}{\sum_i (\mu_{i,\theta})^2} , \tag{21}$$

which is the desired Cramer-Rao bound on the variance of the parameter estimate. The derivative can be evaluated by using the expression for the logistic map, and differentiating to obtain a recursion for the derivative,

$$d_\theta \mu_i = \theta d_\theta \left[\mu_{i-1}(1-\mu_{i-1})\right] + \mu_{i-1}(1-\mu_{i-1}) . \tag{22}$$

Expression (21) for the Cramer-Rao bound therefore roughly means that an efficient estimator (for which eq (18) is an equality) has an error that is inversely proportional to the sensitivity of the mean values of the data points to changes in θ. A plot of the magnitude of the θ derivative of the tenth iterate of the logistic map at $\theta = 3.7$ versus initial conditions is shown in figure 5. For the points where the magnitude of the derivative is large, the Cramer-Rao bound is small. Note that the derivative is also sensitive to the initial condition x_0. This means that for some values of x_0, small changes in θ will not greatly affect the data sequence. At these values it is impossible to obtain a good estimate of the parameter value. Because the bound decreases exponentially with n for a typical trajectory, equation (21) indicates

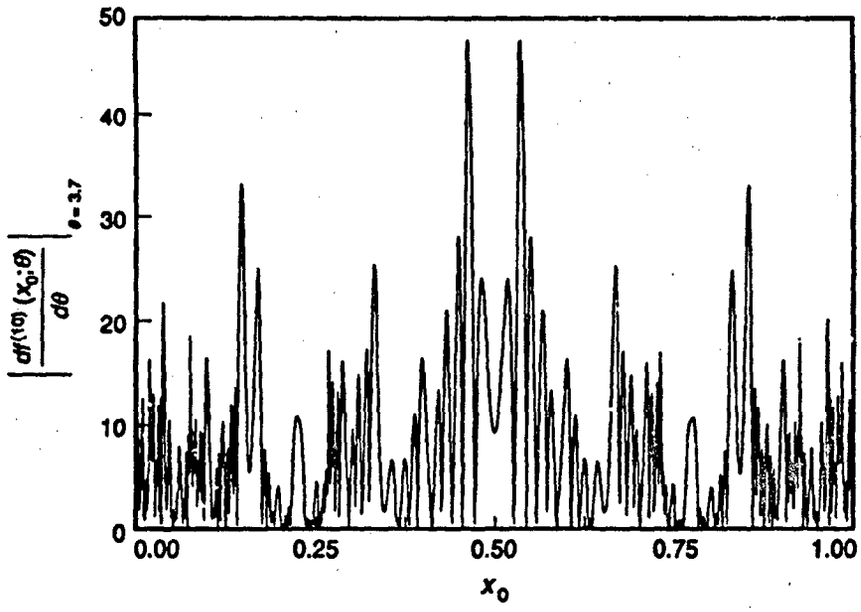


Figure 5. Magnitude of θ derivative of tenth logistic map iterate.

that an efficient estimator would be far superior to those considered in the previous section. The existence of an efficient estimator is now demonstrated, and practical methods for its implementation are given. It is also shown that although uncertainty in the exact values of x limits the accuracy of the estimator in the presence of noise, it is still superior to the ones that have been used before.

6. Optimal Parameter Estimator for One-Dimensional Maps

The Cramer-Rao bound suggests that an estimator for the logistic map may exist such that the error in the estimated parameter decreases exponentially as a function of the number of samples. It is now shown that such an efficient estimator does exist and that it achieves the Cramer-Rao limit, but only for certain chaotic trajectories. Practically, the estimator is better than the usual methods for all data blocks, but the introduction of uncertainties in the mean value of the data vector by additive noise causes the performance to fall below the Cramer-Rao limit for some data vectors.

As a first step in finding the most efficient estimator, assume that the noise variance is small enough so that the map can be linearized in the region of interest. First, for sufficiently small noise amplitudes, consider the estimator formed by a linear combination of estimator functions operating only on pairs of data points at a time:

$$\hat{\theta} = \sum_{i,j} w_{ij} \Theta_{ij}(x_i, x_j) \ . \tag{23}$$

Because any estimator can be expanded in a Taylor series for small noise amplitude as a linear combination of functions of each sample, it can be expanded as functions of data pairs. This expression is therefore general enough to represent any small-noise estimator for the data block. First, consider an estimate of the parameter based on the (x_i,x_j) data pair for a given data sequence. (From this point on consider a hypothetical fixed data sequence and drop the functional notation used in eq (23).) The only estimator that is unbiased is the one that chooses the parameter to be such that $f^{ij} = x_j$ passes through the embedded (x_i,x_j) pair. (This notation means that the $(j-i)^{\text{th}}$ iterate of the map operates on the data point x_i to produce the value x_j.) It is true, however, that there may be multiple values of θ that cause coincidence of this iterate of the map with the embedded data, but it is assumed that the estimate of θ is resolved of ambiguities by successive applications of the technique with lower-order iterates of the map.

If the map is linearized in the region of the embedded data pair, the estimate of θ can be written in terms of derivatives of the map iterates with respect to both x and to θ. The estimate is given by

$$\hat{\theta}_{ij} = \theta - \frac{f^{ij}{}_{,i}\xi_i - \xi_j}{f^{ij}{}_{,\theta}} \quad , \tag{24}$$

where $f^{ij}{}_{,i}$ is the derivative with respect to x_i of the $(j-i)^{\text{th}}$ iterate of the map operating on point x_i, and $f^{ij}{}_{,\theta}$ is the θ derivative. The geometrical meaning of equation (24), as well as some intermediate steps in the derivation, are shown in figure 6. The embedded noiseless pair $(\tilde{x}_i,\tilde{x}_j)$ is perturbed to the observed pair location $(x_i,x_j) = (\tilde{x}_i + \xi_i, \tilde{x}_j + \xi_j)$ by the noise vector (ξ_i,ξ_j) in the function space (function plane) of the $(j-i)^{\text{th}}$ iterate of the map. (The iterated map function f^{ij} passes through the noiseless pair.) Now $\delta\theta \cdot f^{ij}{}_{,\theta}$ is the vertical displacement of the function f^{ij} required for a given small change in θ (given by $\delta\theta$) to make the iterated map pass through the noise-perturbed pair. This displacement is geometrically given by the vertical noise perturbation ξ_j minus the local linear approximation to the vertical displacement caused by the horizontal noise perturbation, which is given by $f^{ij}{}_{,i}\xi_i$. The error in θ caused by the noise is $\delta\theta$; so $\hat{\theta}_{ij} = \theta + \delta\theta$ is the estimate of θ. Using equation (24), the covariance of the pairwise estimates is

$$\text{Cov}\left[\hat{\theta}_{ij},\hat{\theta}_{kl}\right] = \frac{\sigma^2}{f^{ij}{}_{,\theta} f^{kl}{}_{,\theta}}\left[f^{ij}{}_{,i} f^{kl}{}_{,k}\delta_{ik} - f^{ij}{}_{,i}\delta_{il} - f^{kl}{}_{,k}\delta_{jk} + \delta_{jl}\right] \quad . \tag{25}$$

To expedite the derivation, it is now shown that because the map is iterated sequentially to produce the data sequence, any linear combination of three pairwise estimators that are fully linked is degenerate. Fully linked pairwise estimators are defined by their connectivity graph, an example of which is shown in figure 7(a). (Fig. 7(b) and (c) are for later reference.) In this graph,

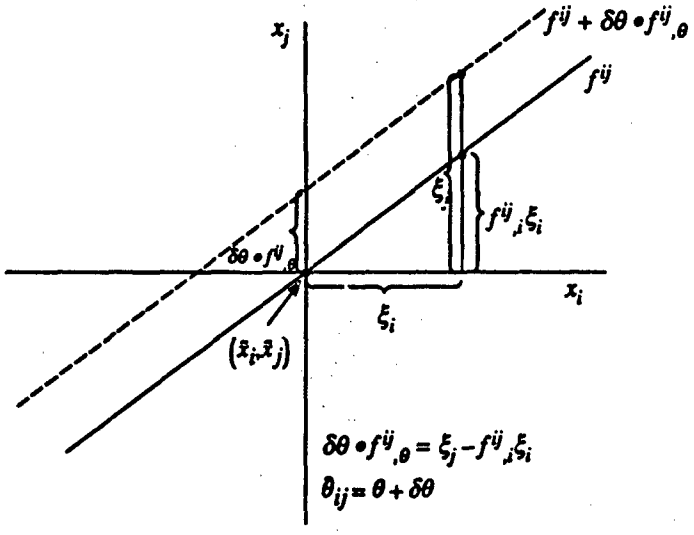


Figure 6. Graphical derivation of pairwise estimator.

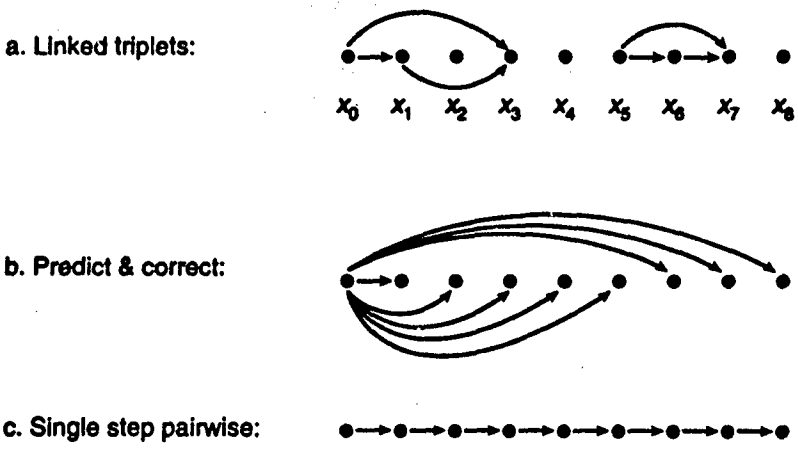


Figure 7. Graphs depicting (a) linked triplets of pairwise estimators, and (b and c) two implementations of the optimal estimator.

an arrow connecting two data points indicates a pairwise estimator operating on the data points. A linear combination of three pairwise estimators such that a complete circuit of all three points can be made is said to be fully linked, or to form a linked triplet. The degeneracy of a linked triplet is readily shown by computing the determinant of the matrix of coefficients of each independent Gaussian random variable in equation (24) for the [(*ij*), (*jl*), (*il*)] triplet of pairwise estimates:

$$\det[L] = \det \begin{bmatrix} \dfrac{f^{ij}{}_{,i}}{f^{ij}{}_{,\theta}} & 0 & \dfrac{f^{ik}{}_{,i}}{f^{ik}{}_{,\theta}} \\ \dfrac{-1}{f^{ij}{}_{,\theta}} & \dfrac{f^{jk}{}_{,j}}{f^{jk}{}_{,\theta}} & 0 \\ 0 & \dfrac{-1}{f^{jk}{}_{,\theta}} & \dfrac{-1}{f^{ik}{}_{,\theta}} \end{bmatrix} . \qquad (26)$$

Expanding the determinant gives

$$\det[L] = \frac{-f^{ij}{}_{,i} f^{jk}{}_{,j}}{f^{ij}{}_{,\theta} f^{jk}{}_{,\theta} f^{il}{}_{,\theta}} + \frac{f^{ik}{}_{,i}}{f^{ij}{}_{,\theta} f^{jk}{}_{,\theta} f^{il}{}_{,\theta}} \tag{27}$$

and using the identity

$$f^{ij}{}_{,i} f^{jk}{}_{,j} = f^{ik}{}_{,i} \tag{28}$$

and equation (27) yields $\det[L] = 0$. The linked triplet is therefore degenerate and can be reduced by deletion of any one pairwise estimator without affecting the quality of the estimate. Continuing in this manner, a reduced set of estimators can be found for any linked set of three or more estimators. This reduced set can be any set of estimators on pairs of data points such that there are no linked triplets, but the set must contain all the sequence points. The estimate for a given data vector can be written as a linear combination over the reduced set as

$$\hat{\theta} = \sum_i w_i \hat{\theta}_i \tag{29}$$

and the variance is minimized by the Lagrange multiplier equation

$$\partial_k \left[\sum_{i,j} C_{ij} w_i w_j - \lambda \sum_i w_i \right] = 0 \ , \tag{30}$$

with the constraint equation ensuring an unbiased estimate given by

$$\sum_i w_i = 1 \ . \tag{31}$$

Here $C = [C_{ij}]$ is the covariance matrix of the pairwise estimators composing the estimator on the full data vector. Equation (30) with constraint equation (31) has the solution

$$w_k = \frac{\sum_l C^{-1}{}_{kl}}{\sum_{k,l} C^{-1}{}_{kl}} \ , \tag{32}$$

and with the use of the constraint equation, the value of the minimal variance is

$$\mathrm{Var}\left[\hat{\theta}\right] = \frac{1}{\sum_{i,j} C^{-1}{}_{ij}} \ , \tag{33}$$

This equation is true only for the true covariance matrix derived above. For other weight coefficients, the variance must be calculated using

$$\mathrm{Var}\left[\hat{\theta}\right] = \sum_{i,j} C_{ij} w_i w_j \quad , \tag{34}$$

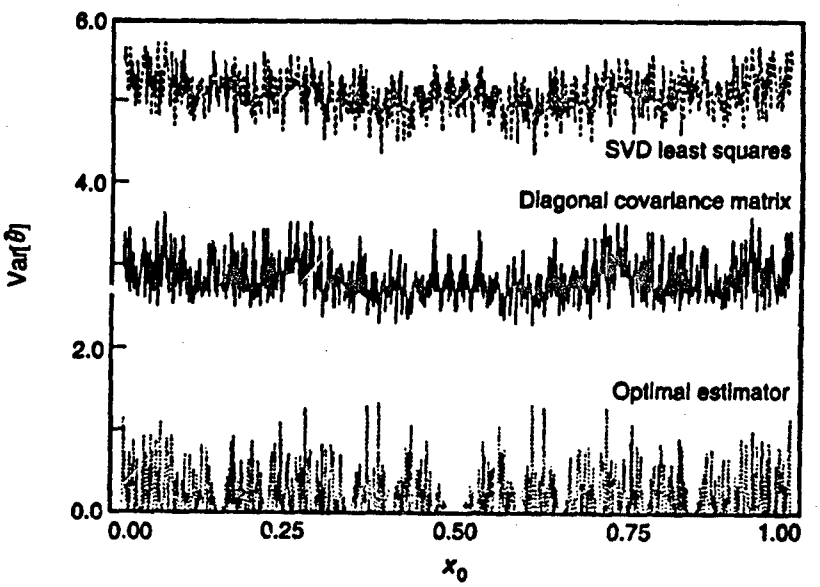
where C is the true covariance matrix.

The theoretical error variance for three different estimators is shown in figure 8. The worst performance (upper curve) is given by the SVD implementation of the least-squares technique, equivalent to the crude estimator considered previously. The middle curve is for an estimator that uses the correct values for the variance of the estimates, but only uses the diagonal form of the covariance matrix in minimizing the variance of the overall estimate. This estimator might be more practical to use in situations prohibiting the inversion of a large and possibly ill-conditioned covariance matrix. The best performance curve (lower curve) is for the optimal technique developed here. The optimal estimator is, however, degraded relative to the Cramer-Rao bound for some trajectories because of the effects of the x derivative term in equation (24). This x derivative term accounts for the uncertainty in the mean data vector, which leads to an uncertainty in the covariance matrix, and thus makes it impossible to attain the Cramer-Rao bound for all data vectors.

7. Implementing the Optimal Estimation Algorithm

The estimator described above attains the Cramer-Rao LMSE bound only for some sequences. Also, if the estimator is used on long sequences, the

Figure 8. Theoretical error variance (performance) for three different estimators.



theoretical performance curves are no longer valid because the linear approximations become inadequate. It is apparent, however, by inspection of the performance curves, that the optimal estimator has several minima, and this property can be used in a more complex estimator that uses the optimal estimator as a block estimator and then combines the blocks to reduce the error. There is also a high probability for a long data sequence that x will eventually fall within the central minimum of the performance curve for a higher order. The parameter can then be detected rapidly with exponentially decreasing error until the width of the minimum becomes too small.

Two possible implementations of the optimal estimator are now outlined:

(1) In the *predict-and-correct* implementation, the successive iterates of the map are used to make finer estimates of the parameter. Because the higher-order iterates of the map have ambiguous values of θ, the earlier values are used to both reduce the variance in the weighted sum and resolve the ambiguities in θ. Figure 9 illustrates the ambiguity problem. This figure shows that for a given value of the tenth iterate of the point $x_0 = 1/2$, it is in general impossible to determine unambiguously the value of the parameter θ, because many values of θ will produce the same value of the tenth iterate. The predict-and-correct implementation has the advantages of functioning in an intuitively obvious manner and being easy to implement. The higher order iterates are used to zero in on the parameter because they are more sensitive to changes in the parameter. Unfortunately, since higher order iterates do not have explicit solutions for the parameter, this estimator uses a linear prediction of θ based on the last estimate for higher orders. This method is illustrated in graph form in figure 7(b). This graph illustrates that the

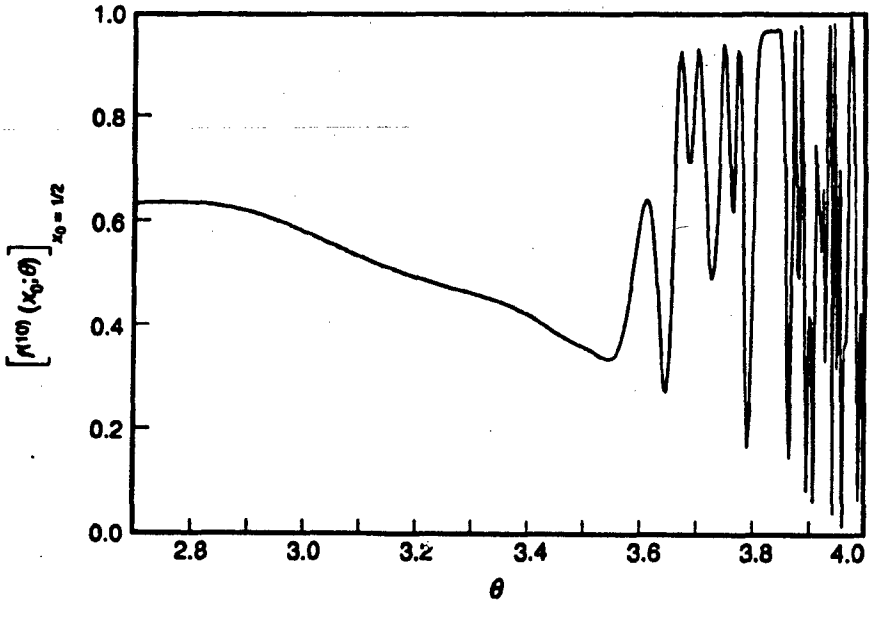


Figure 9. Tenth iterate of $x_0 = 1/2$ versus θ, illustrating ambiguous estimates.

predict-and-correct estimator is constructed by combining pairwise estimators with the same x_0, but using progressively higher iterates of the map.

(2) In the *single-step pairwise* implementation, the degenerate nature of the estimates is taken advantage of to produce a simple implementation without ambiguities. The pairwise estimators used are simply the one-step estimators, and the effect of combining them is to give the linear combination the required sensitivity to changes in θ. Remarkably, the sensitivity-to-parameter effect is reflected in this implementation in the same way as when higher order iterates are directly used as with the predict-and-correct technique. This implementation circumvents the necessity of the linear prediction needed for the predict-and-correct technique, because there is an explicit solution for θ with no ambiguities. This technique is illustrated graphically in figure 7(c).

The single-step pairwise technique has been implemented on the computer and tested using computer-generated noise. The normalized deviation of the parameter error variance for the 10-iterate simulated optimal estimator from the theoretically predicted variance is shown in figure 10. The noise distribution used was uniform, with a peak-to-peak amplitude of 0.001. (The noise variance was therefore 0.001/12.0.) Since the simulated variance was averaged over 1000 trials, the error in the simulation results is negligible. It has been found that for peak-to-peak noise amplitudes of up to about 0.01, this estimator performs with less than twice the predicted error. This deviation is probably entirely due to the local linear approximations to the map equations.

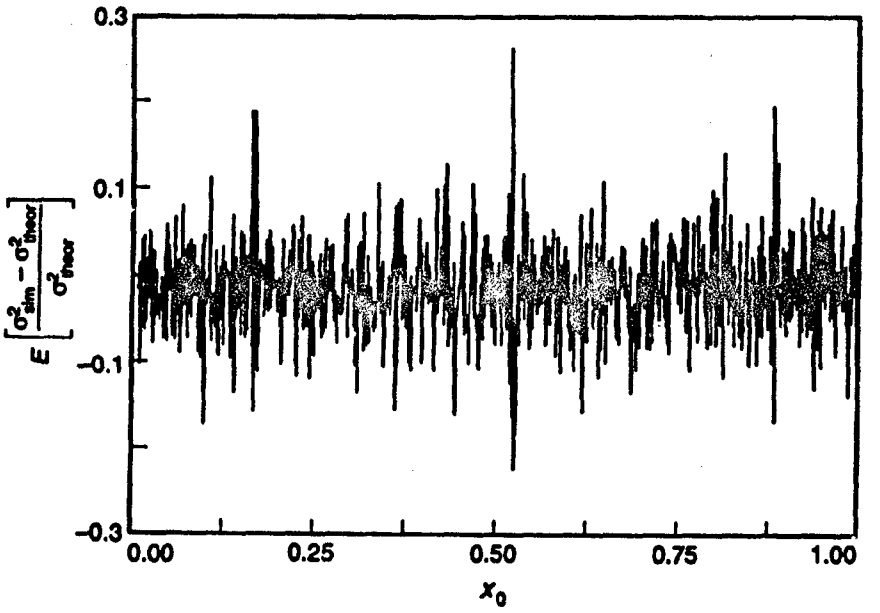


Figure 10. Computer simulation results for optimal estimator.

8. Parameter Detection as Information Acquisition

It is now demonstrated that using the optimal estimation technique for the parameter value, parameter information is extracted at a finite rate. By inspecting the form of the Cramer-Rao bound, if the initial point x_0 is known with great accuracy, the estimate of the parameter has an accuracy that increases exponentially with data sequence length. This behavior is also true if x_0 falls within a minimum in the error curve for the higher iterates of the map. It is shown in this section that, even with a small but finite noise amplitude, this exponential convergence of the estimate to the correct parameter value can be described by a finite parameter information acquisition rate. For simplicity, assume that the initial point is $x_0 = 1/2$, and that the noise amplitude is small. Using only the N^{th} iterate of the map, the variance in the estimate of θ is given by

$$\mathrm{Var}\hat{\theta}_{0N} = \frac{\sigma^2}{\left(f^{0N}{}_{,\theta}\right)^2} \ , \tag{35}$$

where σ^2 is the noise variance. If the (arbitrarily small) probability of error in the k^{th} digit of $\hat{\theta}$ is desired to be p_e^k, then the width of the interval that contains the real value of θ with probability $1 - p_e^k$ is given by

$$\delta\theta|_{p_e^k} = g\left(p_e^k\right)\sqrt{\mathrm{Var}\left[\hat{\theta}_{0N}\right]} = g\left(p_e^k\right)\frac{\sigma}{\left|f^{0N}{}_{,\theta}\right|} \ . \tag{36}$$

Here $g(p_e^k)$ is the number of standard deviations of the density function that must be included in the interval of $1 - p_e^k$ certainty so that the probability of error is p_e^k. The number of digits (in the fractional part of θ) detected at the desired accuracy is thus given by

$$10^{-n_d} = g\left(p_e^k\right)\frac{\sigma}{\left|f^{0N}{}_{,\theta}\right|} \ . \tag{37}$$

Solving for the number of digits resolvable yields

$$n_d = -\log_{10}\left[g\left(p_e^k\right)\frac{\sigma}{\left|f^{0N}{}_{,\theta}\right|}\right] \ . \tag{38}$$

The asymptotic digit reception rate is therefore given by

$$\frac{dn_d}{dN} = \lim_{n\to\infty}\frac{1}{N}\log_{10}\left|f^{0N}{}_{,\theta}\right| \ , \tag{39}$$

which, for a discrete-time sequence, is the number of digits received as a function of time. Now, since the map is chaotic, the derivative with respect to θ is asymptotically exponential in N and can be written as

$$f^{0N}{}_{,\theta} = A \cdot 10^{\alpha N} \quad , \tag{40}$$

where A and α are constants. The limiting value of the digit reception rate, or information rate in digits per sample, is therefore given by

$$\frac{dn_d}{dN} = \alpha \quad . \tag{41}$$

The value of α can be determined from the numerical iteration of the θ derivative map given in equation (22).

Because the optimal estimation technique combines the samples to reduce the statistical variance, the information rate for all samples combined can be expressed as

$$\frac{dn_d}{dN} = \lim_{n\to\infty} \frac{1}{N} \log_{10} \left| \sqrt{\sum_{i=1}^{N} \left(f^{0i}{}_{,\theta} \right)^2} \right| \quad . \tag{42}$$

Of course, for either equation (39) or equation (42) to be strictly true as $n \to \infty$, the noise variance must approach zero. For finite times, however, the estimator will converge exponentially and the information acquisition rate will be finite. Because the digit reception rate is independent of noise amplitude for small enough noise levels, this digit reception rate can be interpreted as detecting the parameter information contained in the data stream through a noisy measurement channel.

The Kolmogorov entropy [6] for a discrete-time dynamical system is the rate at which information must be given to specify successive states of a dynamical system to a given degree of accuracy given the past states of the system. The Kolmogorov entropy for a one-dimensional map [3] is given by

$$K = \lim_{n\to\infty} \frac{1}{N} \log \left| f^{0N}{}_{,0} \right| \quad . \tag{43}$$

The Kolmogorov entropy therefore depends upon the sensitivity of iterates to changes in the initial conditions. A similar function in terms of the parameter θ can be defined as

$$\hat{H}_\theta = \lim_{n\to\infty} \frac{1}{N} \log \left| f^{0N}{}_{,\theta} \right| \quad , \tag{44}$$

and can be interpreted as the additional information about the parameter that can be extracted from the N^{th} iterate of the map given the past. This expression, written by analogy with the expression for the Kolmogorov entropy, is equivalent to the previously developed expression for the information rate. The main difference in the interpretation between the θ entropy and the Kolmogorov entropy is that $\hat{H}_\theta$ refers to the rate at which parameter information is detected, and the Kolmogorov entropy can be interpreted as the rate of refinement of knowledge about initial conditions [7]. Finally, an expression based on the Cramer-Rao bound for the parameter entropy is given by equation (42):

$$H_\theta = \lim_{n\to\infty} \frac{1}{N} \log_{10} \left| \sqrt{\sum_{i=1}^{N} \left(f^{0i}{}_{,\theta} \right)^2} \right| . \tag{45}$$

It is expected that equations (44) and (45) will converge to the same value in the limit as $N\to\infty$ when the map is chaotic, but equation (45) is more meaningful as a finite-time quantity. Finite-time computations of both entropy functions for the logistic map with $\theta = 3.7$ are shown in figure 11. These quantities were computed for the initial condition $x_0 = 1/2$ and for $N = 10$. They can be considered as coarse-grained versions of the limiting values in equations (44) and (45). In the limit as $N\to\infty$, the oscillations of these functions with θ will occur on an arbitrarily small scale, characterizing the extreme sensitivity of the global behavior of the map to the value of θ. In the presence of small amounts of system noise, however, this arbitrarily fine structure will be limited. Also, for chaotic maps, the entropies will become independent of the initial point x_0 for almost all x_0 in the limit as $N\to\infty$. This entropy formalism underscores that for the values of θ for

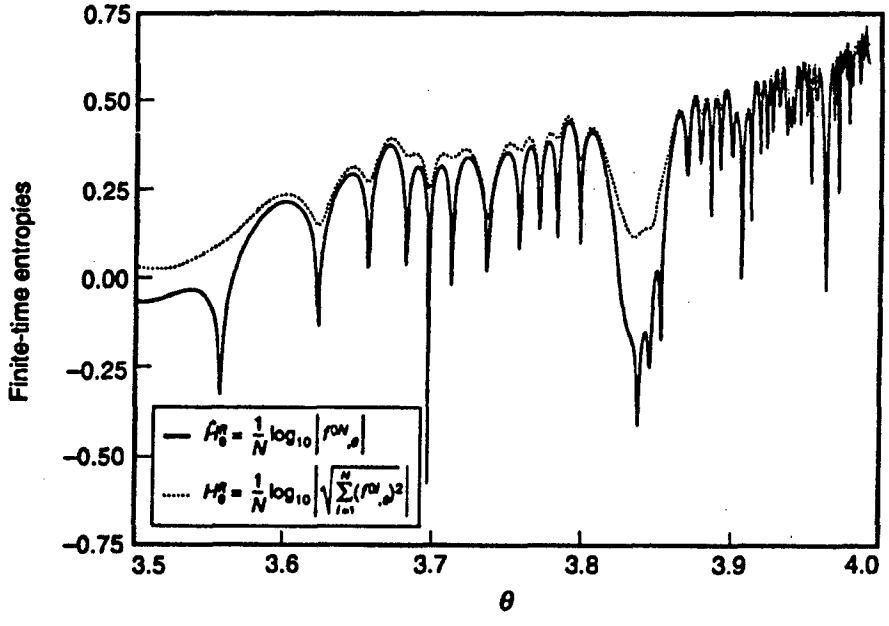


Figure 11. Finite-time parameter entropy functions.

which the map is chaotic, the parameter can be extracted rapidly and accurately, even in the presence of small but finite noise fluctuations. This happens because of the chaotic nature of the data sequence: The sensitivity of the data vector to changes in θ allows one to detect θ accurately, even when the data vector is slightly altered by additive noise. For values of θ where the map is not chaotic, however, the data vector is not as sensitive to changes in θ, and the parameter θ cannot be extracted as rapidly and accurately. Thus, the presence of chaos greatly enhances one's ability to estimate the parameters of iterated maps.

9. Comments and Conclusion

The exponential convergence of the optimal parameter estimator considered here occurs only for certain trajectories of the logistic map. For trajectories that pass close to the point $x_0 = 1/2$ within the length of the data vector, the effect of the x derivative term in equation (24) to reduce the effectiveness of the estimator is not too severe. More precisely, if a data vector is processed using the optimal estimation technique, the estimator will converge exponentially to the correct parameter for m time steps, where m is the largest value of k for which the initial point of the data vector x_0 falls in a minimum of $f^{(k)}(x_0)$. Even here the definition of a minimum of the iterated map is left deliberately vague, because the value of k at which exponential convergence ceases is fuzzy. How close the trajectory must pass to $x_0 = 1/2$ depends upon the length of the data vector used. For $N = 30$, for example, exponential convergence occurs when the initial point x_0 falls near a prominent minimum of the optimal estimator performance curve in figure 8. In the limit, as the length N of the data vector goes to infinity, these minima become smaller until the exponential convergence occurs only in zero noise. (This is because the regions over which the high-order iterates of the logistic map have near-zero slope become very small, and cannot be expected to enhance the estimate in a practical implementation.) Practically, however, for a finite data vector length, it is possible to detect the parameter with exponentially increasing accuracy, even in the presence of finite additive noise for these trajectories. It is also apparent from the performance curve of figure 8 that even when exponential convergence is damped, the estimator converges more rapidly than the expected $1/\sqrt{N}$ convergence for an estimator that relies on convergence of a sum of estimates to the mean.

To illustrate the importance of the values of x_0 with zero slope, the application of the optimal estimation technique is considered for the tent map. The tent map is a symmetric piecewise-linear triangular map with both sides having constant slope, and is described by the equation $x_{n+1} = 1 - 2a|x_n - 1/2|$. There are therefore no values of x_0 for which there is zero slope. For simplicity, consider the variance of an estimate of the gain pa-

rameter a, or equivalently of $b = 2a$, based only on the first and last points in a data vector of length N. (It is assumed that ambiguities have been resolved using lower order pairwise estimates.) The value of the N^{th} iterate of the tent map can be written as $f^{(N)}(x_0) = b^N\Delta$, where Δ is the distance of the initial point x_0 from the nearest point x where the iterated map function $f^{(N)}(x)$ is zero. (The iterated map function consists of 2^{N-1} triangular humps, and thus 2^N line segments pieced together, so Δ is just the distance from the base of the line segment containing the point x_0.) Using equation (24), the variance of an estimate of b can be written as $\sigma^2_{0N} = E[\hat{b}_{0N} - b]^2 = E[-(b^N\xi_i - \xi_j)/(x_0Nb^{N-1})]$, where the initial point has been chosen for convenience to lie on the first line segment of the iterated map function so that $\Delta = x_0$. (This restriction will not affect the N dependence of the computed variance because the magnitude of the slope of the iterated function is constant.) Letting $c = b^2$ for notational clarity and carrying out the calculation for the variance yields $\sigma^2_{0N} = (\sigma^2 c/x_0^2)[(1 + 1/c^N)/N^2]$. Now it is apparent from the form of the term in square brackets that, for fixed x_0, the error (standard deviation) decreases only inversely with N, and not exponentially. This slowed convergence at all x_0 is caused by the constant slope of the tent map, and thus the absence of any regions where the map or its iterates have a near-zero slope. The effect of the x derivative term in equation (24) is therefore severe enough to destroy the exponential convergence rate of the estimator.

The tent map is also useful to illustrate the equivalence of equations (44) and (45) for the parameter acquisition rate. As previously stated, these two expressions are expected to converge to the same value for chaotic maps, and this can be shown to be true analytically for the tent map. Also stated previously, these expressions are strictly true only if the noise variance is zero, so for the purpose of this derivation, a zero noise variance is assumed. (This is usually the case in mathematical derivations of such quantities. Alternatively, if the initial point x_0 is assumed to be known accurately, the assumption of zero noise variance is also rendered unnecessary, because the estimates will converge exponentially for all chaotic trajectories.) Assume again for convenience that the initial point x_0 lies on the first (with leftmost endpoint at $x = 0$) line segment composing the N^{th} iterate of the tent map, so that equation (44) yields $\hat{H}_b = \lim_{n\to\infty}(1/N)\log|Nb^{N-1}x_0|$. Taking the limit yields $\hat{H}_b = \log b$. If equation (45) is used similarly, then the expression $H_b = \lim_{N\to\infty}(1/N)\log(\Sigma_{i=1}^{N}(x_0ib^{i-1})^2)^{1/2}$ is obtained. The summation in the square root can be expressed as $x_0^2\Sigma_{i=1}^{N} i^2c^{i-1} = x_0^2(\partial/\partial c)c(\partial/\partial c)\Sigma_{i=0}^{N}c^i$, so that with the closed-form expression for a finite power series, this summation becomes

$$x_0^2\sum_{i=1}^{N} i^2c^{i-1} = x_0^2\frac{\partial}{\partial c}c\frac{\partial}{\partial c}\left[\frac{c^{N+1}-1}{c-1}\right] = x_0^2\left[\frac{(N+1)^2c^N}{c-1} - \frac{(2N+3)c^{N+1}-1}{(c-1)^2} + \frac{2\left(c^{N+2}-c\right)}{(c-1)^3}\right]$$

The first term in square brackets dominates for large N, so the expression for the entropy becomes $H_b = \lim_{N\to\infty} (1/N)\log [x_0(N+1)b^N/(b^2-1)^{1/2}]$, so that in the limit $H_b = \log b$. Equations (44) and (45) thus yield identical expressions for the parameter entropy in the limit of large N for the tent map. A direct calculation of these two quantities is possible for the tent map because it is piecewise linear, but the equivalence of the two equations is caused by the exponential increase in the sensitivity of the iterates to changes in the parameter, so that equations (44) and (45) should yield equivalent values for the *chaotic* trajectories of almost all maps.

References

[1] E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).

[2] T. Shinbrot, E. Ott, C. Grebogi, J. A. Yorke, Phys. Rev. Lett. **65**, 3215 (1990).

[3] Heinz Georg Schuster, *Deterministic Chaos*, Weinheim (Federal Republic of Germany): VCH Verlagsgesellschaft (1989).

[4] Mitchell J. Feigenbaum, J. Stat. Phys. **19**, 25 (1978).

[5] Richard E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley (1988).

[6] A. N. Kolmogorov, Dokl. Akad. Nauk **119**, 754 (1958).

[7] J. Doyne Farmer, Z. Naturforsch **37a**, 1304 (1982).

Distribution

Administrator
Defense Technical Information Center
Attn: DTIC-DDA (2 copies)
Cameron Station, Building 5
Alexandria, VA 22304-6145

US Army Missile Command
Attn: AMSMI-RD-WS, C. Bowden
Redstone Arsenal, AL 35898-5248

US Army Research Office
Attn: J. Mink (2 copies)
PO Box 12211
Research Triangle Park, NJ 27709

Naval Research Laboratory
Attn: Code 5540, J. Palmer
4555 Overlook Ave., SW
Washington, DC 20375-5336
Attn: Code 5540, S. Lessin

Clemson University
Electrical Engineering Department
Attn: C. Butler
Clemson, SC 29634

George Washington University
Department of Electrical Engineering
Attn: P. Kakaes
Washington, DC 20052

Northwestern University, Laboratory for Chaos, Dept of Chemical Engineering
Attn: T. Shinbrot
Evanston, IL 60208

Rensselaer Polytechnic Inst, Dept of Physics
Attn: J. Haus
Attn: R. Lichtenstein
Troy, NY 12180-3590

University of Maryland
Department of Physics and Astronomy
Attn: D. Currie
College Park, MD 20742

University of Maryland
Laboratory for Plasma Research
Attn: C. Grebogi
Attn: E. Ott
College Park, MD 20742

U.S. Army Research Laboratory
Attn: AMSRL-OP-CI-AD, Mail & Records Mgmt
Attn: AMSRL-OP-CI-AD, Tech Pub (2 copies)
Attn: AMSRL-OP-CI-AD, Library (3 copies)
Attn: AMSRL-E, D. Wortman
Attn: AMSRL-E, M. Tobin
Attn: AMSRL-SL-ND, C. Glenn
Attn: AMSRL-SL-ND, M. Smith
Attn: AMSRL-SS, D. Rodkey
Attn: AMSRL-SS-FG, R. Tobin
Attn: AMSRL-SS-FS, J. Sattler
Attn: AMSRL-SS-M, N. Berg
Attn: AMSRL-SS-M, A. Sindoris
Attn: AMSRL-SS-M, B. Weber
Attn: AMSRL-SS-S, V. DeMonte
Attn: AMSRL-SS-S, J. Miller
Attn: AMSRL-SS-SF, J. Goff
Attn: AMSRL-SS-SF, N. Gupta
Attn: AMSRL-SS-SF, J. Pellegrino
Attn: AMSRL-SS-SF, B. Sadler
Attn: AMSRL-SS-SG, J. Sichina
Attn: AMSRL-SS-SI, A. Filipov
Attn: AMSRL-SS-SI, M. Patterson
Attn: AMSRL-SS-SJ, B. Stann
Attn: AMSRL-WT, D. Eccleshall
Attn: AMSRL-WT, J. Frasier
Attn: AMSRL-WT-N, Director
Attn: AMSRL-WT-N, J. M. McGarrity
Attn: AMSRL-WT-NB, H. Brandt
Attn: AMSRL-WT-NB, J. Soln
Attn: AMSRL-WT-NE, C. Fazi
Attn: AMSRL-WT-NG, A. Bromborsky
Attn: AMSRL-WT-NH, J. Corrigan
Attn: AMSRL-WT-NH, R. Gilbert

Distribution (cont'd)

US Army Research Laboratory (cont'd)
Attn: AMSRL-WT-NH, S. Hayes (30 copies)
Attn: AMSRL-WT-NH, G. Huttlin
Attn: AMSRL-WT-NH, L. Libelo
Attn: AMSRL-WT-NH, M. Litz

US Army Research Laboratory (cont'd)
Attn: AMSRL-WT-NH, B. Ruth
Attn: AMSRL-WT-NI, A. Abou-Auf
Attn: AMSRL-WT-PC, A. Cohen
Attn: AMSRL-WT-PC, R. Beyer